E-Print Archives: A Fast Emerging Mode Of Electronic Communication Of Pre-Prints

E. R. Prakasan, Anil Kumar, Anil Sagar, Lalit Mohan, V. L. Kalyane, Vijai Kumar and Rekha P. Upadhye

Library and Information Services Division, Bhabha Atomic Research Centre, Trombay, Mumbai– 400 085, India E-Mail: prak@magnum.barc.ernet.in

ABSTRACT

This study is based on a survey of discipline-based preprint archives available in electronic format on the Internet. The process of peer-review in the e-print (preprints in electronic format) age is presented. Most popular 19 e-print archive services in different disciplines have been identified. Recent initiatives on e-print archives like OAI, OpCit etc. are noted.

0. Introduction

One of the most prevalent informal forms of exchange of research data and results among scientists was personal correspondence through letters until the 17th century [Kalyane and Kalyane, 1994]. Again communication took many forms, both formal and informal, and has evolved over the centuries.

The term "preprint" most often refers to a manuscript that has gone through a peer-review process and now awaits publication in a traditional journal. A preprint accessible over the Web may also be referred to as an "e-print." Preprints also cover papers that authors have submitted for journal publication, but for which no publication decision has been reached, or even papers electronically posted for peer consideration and comment before submission for publication. In fact, preprints can also be documents that have not been submitted to any journal.

The American Physical Society expands this, indicating that the concept of e-prints includes any electronic work circulated by the author outside of the traditional publishing

environment [American Physical Society]. In some subjects, where rapid transmission of knowledge is critical, electronic dissemination of preprints is an absolute necessity, with subsequent traditional publication becoming almost a formality. In mathematics and physics, for example, formal publication provides archiving, which serves more to remind the scholarly community of the paper's initial appearance. Ultimately, formal publication serves as a vehicle to support the standing of the author [Lim].

D. Lim (1996) defines preprints as manuscripts, which may fall into one or more of the three categories:

- that have been reviewed and accepted for publication,
- that have been submitted for publication but for which a decision to publish has not been made, and
- that are intended for publication but are being circulated among peers for comment prior to being submitted for publication.

Of particular interest from the perspective of electronic preprints are those preprints that are intended for publication, and are self-archived by their authors either prior to, or after, acceptance for publication. However, 'e-print' (or 'eprint') can be defined in other ways. Here, the focus will be on 'e-preprints', with a particular emphasis on those that have been self-archived by their authors.

The name derives from the original meaning: a preprint (pre-published article) in electronic form. The concept is now more nebulous, and is general enough to include any electronic work circulated by the author outside of the traditional publishing environment. Here it means any electronic (not necessarily printable), research-related information provided by the author.

Electronic communication provides a potential solution, which enhances speed and accessibility while adding value in a number of ways, although its economics remains unclear. What is certain is that a new pattern for the communication of scientific information will require changes in the pattern of behaviour of all those involved; the author and his or her funding agencies, the facilitators such as publishers and librarians and the end user [Elliott].

A decade ago scholarly communication involved mail, fax, or more recently, FTP, and electronic mail. While traditional production and publication of documents requires a significant investment of time, materials, and money, placing a preprint or e-print on the World Wide Web involves no printing costs and practically no distribution costs.

Electronic communication has created new ways to distribute such results and is forcing researchers and publishers to reassess the old procedures and consider new possibilities as we learn to use the Internet. Now, not only can authors easily disseminate their results, but also networked readers can have cheap, fast access to more scientific literature and have it in a form that facilitates its use in their own research [Bachrach et al, 1998]

Scholarly communication converges on a stable set of electronic forums, such as "pre-print" servers, discussion lists, and electronic journals. A second common argument is that the variety of e-media initiatives reflects a creative period in scientific communication, and this, in itself, is a good thing [Kling and McKim, 2000].

With on-line archives (both preprints and post-prints), all papers can be located by anyone quickly and easily -- and at no cost. Authors can put draft copies and successive updates up for public view, until the final, peer-reviewed (published) version appears. Users can follow the research through all of its successive stages from pre-prints through to the post-prints.

Paul Ginsparg, a physicist at the Los Alamos National Laboratory, developed the first preprint archive in August 1991. Originally dedicated to papers in high-energy theoretical physics, the "arXiv.org e-Print archive" at http://www.arxiv.org/ took several months to attract 1,000 users; presently it reports from 35,000 to 150,000 visits per [http://www.eprints.org/results/report.html]. Professional societies, government sites, and universities usually provide preprint services.

Particular questions that were asked in the early stages of the project were:

- Why do authors use the archive?
  Do researchers archive or use the archive more than others?

- At what point does an author decide to archive a draft?
- Do authors cite pre-prints, published post-prints, or both, and under what conditions?
- When an author cites a pre-print, do they update the citation when the cited paper is revised or published?
- If a paper is eventually accepted by a journal, does the author update the text of the paper or just the reference information?
- What are the authors' practices in archiving successive drafts?
- What are the authors' practices in citing successive drafts?
- What is the relationship between the impact factor of an author and download frequency and other online performance indicators and practices?

1. Advantages

The Internet is a boon to the scientific community, students, and the public. Anyone can share information, ideas, and events with others. Electronic preprints need not simply represent what would appear in print journals. E-prints may offer numerous value-added elements, including audio and video, as well as linked references to other documents

E-print archives services are very formal mode of communication in which each entry is archived and indexed for retrieval at later times; Usenet newsgroups and bulletin boards, on the other hand, represent an informal mode of communication, more akin to ordinary conversation, with un-indexed entries that typically disappear after a short time [Ginsparg, 1996]. The time elapsed for getting comments and suggestions from experts working in the same area of work has been drastically reduced. Aim of Libraries will not be to have up-to-date information but up-to-the-minute information [Jones, 991]. As libraries evolve in the changing electronic revolution, they can continue to have an important role in supporting research [Luce]. The peer-review process at this e-print age is depicted in Figure 1. Discussion group interaction has often led to immediate reposting of reworked preprints with acknowledgements to individuals who have offered comments and criticism on the original preprints [Boyce, 2000]

2. Discipline-Wise E-Print Archive Services

Because of the general dishevelment of Web-based information, the biggest problem with exploiting the potential of preprint servers simply knows which servers to use for

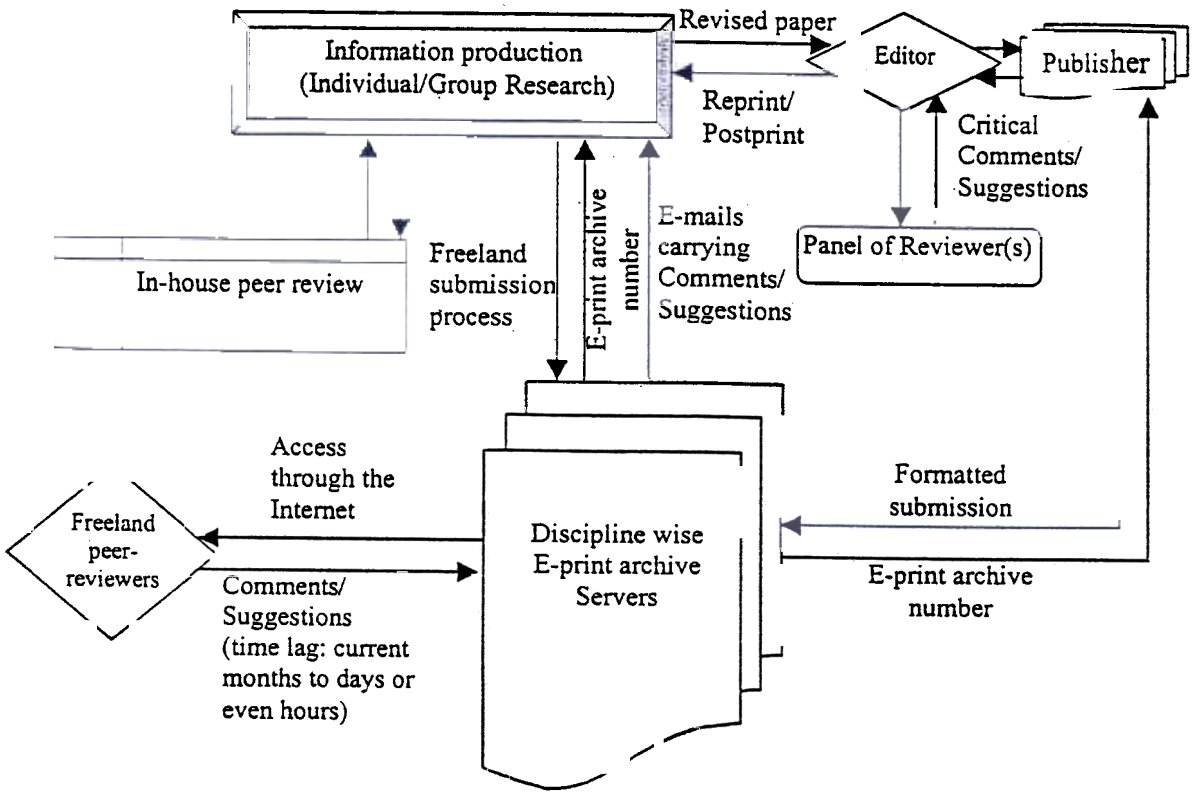

Figure Review of scholarly communications in e-print age

specific queries. Informally shared information among researchers probably accounts for end-user scholars deciding which preprint server to access for specific information. But while subject experts know which specific preprint server to explore, general information specialists may not possess this inside information.

Disciplines such as astronomy, chemistry, computer science, mathematics, and physics have taken the lead in preprint distribution. Perhaps because scientists and researchers in these fields possessed the first high-level computers, preprint servers became available and then prevalent in these disciplines. Fields in the humanities and social sciences have recently begun following the trend, but still lag significantly behind in terms of server repositories and papers.

Study of content analysis of these e-print archives and the bibliometric analysis are very interesting [Prakasan et.al, 2003]. Indeed, for graphically dependent sciences, preprint publication on the Web is preferable to paper journals because of the possibilities for inclusion of audio and video and other multi-dimensional intense graphics. Electronic preprints do not represent the only example of the technological impact of the greater efficiency and storage capacity of digital media. Glimpses of popular discipline based e-print archive services are given in Table 1.

Table 1: Discipline-wise notable preprint sites

Name of the E-print Archive Service	URL	Initiated year	Discipline(s)
LANL's E-Print Archive	http://xxx.lanl.gov/	1991	Physics, mathematics, nonlinear science, and computer science
CERN Document Server: Preprints	http://preprints.cern.ch/	1994	Physics
SLAC SPIRES-HEP	http://www-slac.slac.stanford.edu/find/spires.html	1994	High Energy Physics
Topology Atlas Preprints	http://at.yorku.ca/topology/preprint.htm	1995	Topology
American Physical Society E-Prints	http://publish.aps.org/eprint	1996	Physics
Chemical Physics Preprint Database	http://www.chem.brown.edu/chem-ph.html	1996	Theoretical chemistry
CogPrints	http://cogprints.soton.ac.uk	1997	Psychology, anthropology, philosophy, and linguistics
HUM-MOLGEN	http://www.informatik.uni-rostock.de/HUM-MOLGEN/	1998	Human geneticists and molecular biology
Archives of European integration	http://eiop.or.at/erpa/	1998	European-integration studies
Clinmed Netprints	http://clinmed.netprints.org/	1999	Clinical medicine and health
WoPEc	http://netec.mcc.ac.uk/WoPEc.html	1999	Economics
ChemWeb	http://www.chemweb.com/	2000	Chemistry
MSRI	http://www.msri.org/publications/preprints	2001	Mathematics
E-Math	http://www.ams.org/preprints	Unknown	Mathematics

NCSTRL (Networked Computer Science Technical Reference Library)	http://www.ncstrl.org/	Unknown	Computer science
PrePrint Network Energy)	http://www.osti.gov/preprint	Unknown	Physics, materials, and chemistry, portions of biology, environmental sciences and nuclear medicine.
Social Science Research Network	http://www.ssrn.com/	Unknown	Social Science
Theoretical Ecology Preprint Database	http://www.nceas.ucsb.edu:8504/esa/ppr/ppr.Query	Unknown	Theoretical Ecology
Universal Preprint Directory	http://www.realsci.com/browse.cfm	Unknown	Religion and other philosophy topics

Although some might consider tracking preprints to be the proper responsibility of indexing and abstracting services, for the foreseeable future, Web searching will remain the primary strategy for locating these reports. However, some indexing and abstracting services have finally recognized the value of tracking preprints.

CERN Document Server (CDS) at http://weblib.cern.ch/Home/, One-Shot World-Wide Preprints Search at http://www.ictp.trieste.it/indexes/preprints.html etc. allows category searching for various disciplines [Tomaiuolo, 2000].

3. Preprint Usage and Citations of E-Print Archives

Various reports from the information science literature and other fields indicate that use of preprint servers, especially in the sciences, is very high. Youngen observed that the number of electronically posted preprints in astrophysics doubled every year during the 1992 through 1997 study period. Using the Institute for Scientific Information's SciSearch, slightly over 100 citations to preprints were retrieved in 1995; in 1997, the number of citations to electronic preprints rose to over 400. Youngen concluded, "The growth rate in citations reflects not only the authors' acceptance of the e-print, but the publishers and editors (acceptance) of the manuscripts as well" [Youngen, 1998]. The use of electronic forms of scholarly information has typically been growing at 50 to 100 per cent per year [Odlyzko, 2000].

4. Open Archives Initiative (OAI)

In October, 1999, a meeting was held in Santa Fe, New Mexico, where participants included librarians, publishers, and computer scientists. The unifying goal was the establishment of a universal preprint archive. Laying the foundation for the resolution of technical challenges such as archive maintenance, accessibility, and interoperability, the project was called the "Open Archives Initiative" [Van de Sompel and Carl, 2000]. The Open Archives Initiative activities centered at Cornell University, develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content.

5. Open Citation Linking Project (OPCIT)

The OpCit is a funded project, currently developing tools to make the existing resources more powerful by completely citation inter-linking all of the papers in The Los Alamos E-print Archive (arXiv) and eventually to extend this to all the rest of the disciplines in other open archives. The user need only click on the citation to view that paper -- as long as it too is archived online. For the purposes of the third year project this investigation has been extended to investigate a previously unexplored avenue. A fundamental part of the usage of an on-line archive is the habits of the users themselves. The aim has been to extend the ongoing research by investigating the relation between the objective online indicators and the authors' own verbal reports of their practices and rationale in archiving their work [http://www.eprints.org/results/report.html].

6. Current Frequently Asked Questions (FAQ) And Discussion

- Whether preprints should accept or not for print publication after electronic posting?
  The threat commercial journal publishers feel electronic pre-prints represent to this revenue, and discussion about government involvement in e-print publishing. In a word, it's all about money.
- "Durability" or the issue of archival survival is yet another misgiving.
- Issues like Plagiarism and Copyright.
  Metadata for e-print archives.

7. Conclusion

E-Print archives are still in their infancy, having only started to appear ten years ago. Yet despite the explosion of the WWW, only a minority of authors (30-40 %), and in very few fields (mostly Physics, Mathematics, and Computer Science) is using this channel for their scholarly communication.

It is perhaps the best-known example of the way the Internet has already changed the scholarly communication habits of scientists. The tighter integration of formal and newly rising informal e-print systems represents an enormous opportunity for libraries and information providers all to the benefit of the researcher.

8. References

. American Physical Society (1998). What Are Eprints? [http://publish.aps.org/eprint/docs/faq.html] 25th July, 2000.
2. Bachrach, S., et al, (1998). Who should own scientific papers? *Science*, **281**(5382), pp. 1459-1460.
3. Boyce, Peter B. (2000). For Better or Worse: Preprint Servers Are Here to Stay. *College and Research Libraries News*, **61**(5), pp. 404-407, 414.
4. Elliott, R. The opportunity and challenge of new communications technologies. [http://www.nature.com/wcs/c05.html] 7th Oct., 2003.
5. Ginsparg, P. (1996). Winners and Losers in the Global Research Village. Joint ICSU Press/UNESCO Expert Conference on ELECTRONIC PUBLISHING IN SCIENCE UNESCO, Paris, 19-23 February 1996 [http://gateway.library.uiuc.edu/icsu/ginsparg.htm] 7th Oct., 2003.
6. Jones, K. P. (1991) Information and the information professional. *Aslib Proceedings*, **43**(9), pp. 259-269.
7. Kalyane, V.L. and Kalyane, S.V. (1994) R&D communication strategy vis-à-vis librarianship. *Journal of Information Science*, **4**(3), pp. 105-135.
8. Kling, Rob and McKim, Geoffrey (2000). Not just a matter of time: Field differences and shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science*, 51(14): pp. 1306-1320 [http://www.slis.indiana.edu/CSI/WP/wp99_02B.html] 7th Oct., 2003.
9. Lim, D. (1996). Preprint servers: a new model for scholarly publishing?. *Australian Academic and Research Libraries*, **27**(1), pp. 21-30.
10. Lim, Edward (1996). Preprint Servers: A New Model for Scholarly Publishing? *Australian Academic and Research Libraries*, **27**(1), pp. 21-30.
11. Luce, Richard E. E-prints Intersect the Digital Library: Inside the Los Alamos arXiv [http://www.library.ucsb.edu/istl/01-winter/article3.html] 30th Sept., 2003.
12. Odlyzko, A.M. (2000). The rapid evolution of scholarly communication [http://www.research.att.com/~amo] 7th Oct., 2003.
13. Prakasan, E. R.; Anil Kumar; Anil Sagar; Lalit Mohan; Singh, Sanjay Kumar; Kalyane, V. L. and Vijai Kumar (2003) Content analysis of E-Print Archives. http://arxiv.org/abs/physics/0308107

14. Tomaiuolo, Nicholas G. (2000). Preprint Servers: Pushing the Envelope of Electronic Scholarly Publishing. Searcher, **8**(9) [http://www.infotoday.com/searcher/oct00/tomaiuolo&packer.htm] 7th Oct., 2003.
15. Van de Sompel, Herbert and Lagoze, Carl (2000). The Santa Fe Convention of the Open Archives Initiative. *D-Lib Magazine*, **6**(2) [http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html] 30th Sept., 2003.
16. Youngen, Gregory K. (1998). Citation Patterns to Electronic Preprints in the Astronomy and Astrophysics Literature. *Library and Information Services in Astronomy*, **153**.